(2)

RSRE
MEMORANDUM No. 4511

ROYAL SIGNALS & RADAR ESTABLISHMENT

PRELIMINARY RESULTS ON THE USE OF LINEAR DISCRIMINANT ANALYSIS IN THE *ARM* CONTINUOUS SPEECH RECOGNITION SYSTEM

Authors: S M Peeling & K M Ponting

PROCUREMENT EXECUTIVE,
MINISTRY OF DEFENCE,
RSRE MALVERN,
WORCS.

RSRE MEMORANDUM No. 4511

UNLIMITED

92 3 16 108

CONDITIONS OF RELEASE

0120199 308899

************************* DRIC U

************************* DRIC Y

Reports quoted are not necessarily available to members of the public or to commercial organisations.

Royal Signals and Radar Establishment

Memorandum 4511

Preliminary Results on the Use of Linear Discriminant Analysis in the *ARM* Continuous Speech Recognition System

S M Peeling and K M Ponting

16th December 1991

Abstract

Linear discriminant analysis is used to generate speech data transformations. The transformed data is then used within the *ARM* continuous speech recognition system. Preliminary results are presented from experiments using transformed data alone and also in conjunction with one, or both, of word transition penalties and variable frame rate analysis. Speaker dependent results are reported which are significantly better than the best obtained previously.

INTENTIONALLY BLANK

Contents

List of Figures

List of Tables

1 Introduction

The work described in this report was conducted at the UK Speech Research Unit. It is partly supported by IED project 3/1/1057 on Speech Recognition Techniques and also forms part of the Airborne Reconnaissance Mission (*ARM*) continuous speech recognition project. The aim of the *ARM* project is accurate recognition of continuously spoken airborne reconnaissance reports using a speech recognition system based on phoneme-level hidden Markov models (HMM). The *ARM* project is described in detail in [15].

The *ARM* system currently applies a discrete cosine transformation to a spectral representation of the spe:ch to produce (so called) mel frequency cepstral coefficients (MFCCs). This linear transformation and representation is commonly used in current speech recognition systems (eg [6], [8]).

Linear discriminant analysis (LDA) can be used to transform data in order to improve a classification system and has the advantage of determining the relative importance of the transformed coefficients in the discrimination process. This allows for some degree of (informed) data reduction. A fuller description of LDA can be found in Section 2.

The LDA transformation has been applied to speaker dependent data in the *ARM* system. Previous papers have shown that the performance of the *ARM* system can be improved by using VFR analysis and word transition penalties to reduce the numbers of insertions (eg [13]). Results are presented here using the LDA transformation on its own (Sections 4.1.1 and 4.2.1), with word transition penalties (Sections 4.1.2 and 4.2.2) and in combination with VFR analysis (Sections 4.1.3 and 4.2.3).

2 Linear Discriminant Analysis

This section will give a broad overview of LDA; for a more detailed description see [4], [5].

In any pattern classification task the main objective is to assign some unknown pattern to a particular class. In order to achieve this, it is necessary to attempt to match one set of features against another. Ideally this set of features should not be too large and there should be some information as to the relative importance of individual features in the classification process.

In speech recognition the cosine transformation is commonly used to improve the discrimination process (and to reduce the number of features in some systems). One motivation for the use of this transformation was given by Pols ([11]). He showed that the first three cosine components were a reasonable approximation to the first three principal components of his speech data.

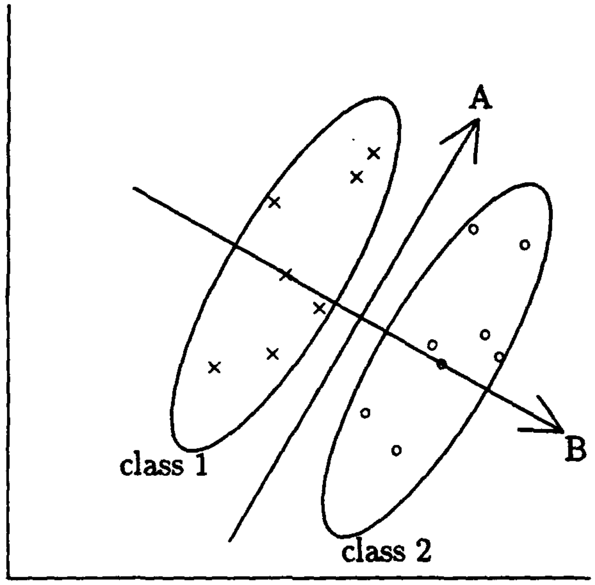

Figure 1: The first two principal components for the two classes of (artificial) data.

However principal components analysis is primarily concerned with the total covariance matrix of the input data and takes no account of any known class labels. Therefore the *improvement in discrimination* is a by-product of this analysis, rather than its chief aim. This can be seen from the artificial data shown in Figure 1. Principal components analysis will give direction A as the first principal component, and B as the second, but all discrimination relies on B.

Linear discriminant analysis provides a method of examining class-labelled data and discovering a set of features which are important in the discrimination process. LDA has the added advantage that these features are ordered so that their relative importance in this discrimination process is known. Because of this, LDA can be used to provide a reliable means of data reduction.

It is worth noting that LDA applied to the data in Figure 1 would give direction B as the first linear discriminant.

Geometrically, the LDA transformation corresponds to a rotation followed by a scaling followed by a rotation of n dimensional space. These are constructed so that variations between the classes are concentrated in the lower dimensions of the space.

The LDA analysis assumes that the within class covariance matrix is the same for each class and relies only on pooled within (W) and between (B) class covariance matrices. After the transformation, the corresponding W' is the identity matrix and B' is diagonal with the variances down the diagonal ordered by size. This means that the set of features which give the greatest between class discrimination can easily be extracted (ie the less important features can be discarded).

3 Experimental Setup

In all the experiments reported here, the data created was passed to the *ARM* system which is described in [14], [15]. The version of the *ARM* system used here was a triphone based HMM system.

The speech data used were obtained by passing digitised speech signals through a 27 channel filter bank analyser at 100 frames per second. The filters were spaced on a non-linear frequency scale based on that in [3]. As with the experiments reported in [9], the bottom (DC-60Hz) channel was omitted. Hence only the top 26 channels output from the filter bank were used.

The class labels used for LDA were based on forced alignment of the training data to previously generated HMMs. Each speech frame was given a class label indicating the phoneme and model state within the aligned triphone models. Hence, since most of the models contained three states, there were three classes for each phoneme. These class labels were then used to calculate pooled within class and total covariance matrices; the between class matrix being obtained by subtraction.

Many different transformations can be obtained by using different representations of the filter bank speech data. The simplest is to consider a single frame of data and its associated class. However it can be useful to include information from surrounding frames. For example, three input frames at a time could be considered, with the relevant class being determined by the centre frame. In more complicated schemes differences can be incorporated whereby instead of considering surrounding frames directly, the differences between them are used. Similarly, regression coefficients over several frames can be used. In [5], Hunt and Lefébvre report using log filter outputs, regression coefficients and a notch filter representation as the primary representation to which LDA is applied.

In the experiments reported here, two different schemes for calculating the LDA transformation have been employed. In the first, and simplest case, the analysis considered a single frame at a time. This will be referred to as a "single frame transform" and the transform matrices created in this way thus contained 26 × 26 elements. In the second case, three frames of input were used, with the classification dependent upon the central frame. This is referred to as a "three frame transform" and the matrices contained 78 × 78 elements.

Clearly in the case of the three frame transform, it was not practical to use the complete output vector. Since LDA had ordered these elements it was to be expected that some of them could be discarded. Experiments were conducted to gain an insight into how many of the elements were needed and the results are reported in Section 4.1.1, for the single frame transform, and Section 4.2.1, for the three frame transform.

It was shown in [13] that the recognition performance of the *ARM* system could be significantly improved by the use of word transition penalties which were used to control the relative numbers of insertions and deletions. Results are reported here for a range of word transition penalties.

Experiments were also conducted into the effect of employing VFR analysis as a further method of data reduction, after the LDA transformation. A full description of VFR analysis can be found in [9]. It is sufficient here to state that VFR analysis is a data dependent method of data reduction. In the VFR experiments, various thresholds were used whilst the duplication limit remained at 50.

Speaker dependent recognition experiments were conducted using speech from two male speakers (namely RKM and MJR) as training and test material. The training set consisted of 37 *ARM* reports per speaker, (224 sentences, 1985 words per speaker) chosen to give maximum coverage of phonemes which occur infrequently in the *ARM* vocabulary. Ten different reports from the same speakers (540 words, 2293 phonemes per speaker) were used as test material.

Recognition was performed using a one-pass dynamic programming algorithm with beam search and partial traceback [1]. Results are presented in terms of *% words (or phonemes) wrong* and *% word (or phoneme) errors*[1]. These are computed as follows, using dynamic programming to align the true transcription of the test data with the output of the recogniser:

$$\% \, words \, wrong = \frac{S+D}{N} \times 100,$$
$$\% \, word \, errors = \frac{S+D+I}{N} \times 100$$

where N is the number of words in the test set, and S, D and I are the number of words recognised as the incorrect word, deleted and inserted respectively.

Recognition results are reported for two levels of syntactic constraint. All the phoneme results come from employing the *phoneme* syntax in which any sequence of triphones can be recognised and the results are scored according to whether or not the correct phoneme is recognised. The word results are obtained from the *word*

[1] Previous papers have quoted percentage word accuracy results which are defined as:–
$100 - \% word \, errors$.

syntax which allows recognition of any sequence of non-speech sounds and words from the *ARM* vocabulary.

Significance levels for the results presented here are obtained using the matched pairs test suggested in [2] and implemented as described in [7].

4 Results

The results are presented in two sections, the first deals with results obtained using single frame transforms whilst the second set used three frame transforms. In both cases, results are reported for various numbers of discarded elements, then also with the addition of word transition penalties and VFR analysis.

4.1 Single Frame Transform

4.1.1 Varying the Numbers of Discarded/Retained Elements

Speaker	No of Elements	Phone		Word	
	Discarded/Retained	Wrong	Errors	Wrong	Errors
MJR	0/26	15.1	38.5	7.2	15.4
	5/21	15.0	40.3	7.0	13.5
	8/18	16.4	43.4	6.7	13.3
	10/16	15.3	44.7	6.3	14.1
	12/14	15.7	48.2	6.5	13.3
	15/11	18.3	61.1	5.7	13.0
	20/6	34.9	126.4	7.8	18.5
RKM	0/26	18.9	41.5	6.1	15.4
	5/21	19.9	44.2	5.4	12.6
	8/18	20.1	46.9	5.7	12.8
	10/16	21.2	51.4	5.4	11.7
	12/14	22.3	54.5	5.7	12.4
	15/11	23.0	62.2	6.3	14.6
	20/6	34.8	107.7	9.3	20.0

Table 1: Full recognition results obtained using single frame transform matrices with various numbers of elements discarded/retained.

Each transformed data frame contained 26 elements and initial experiments were conducted to investigate how many of these elements were important in the discrimination process. Table 1 shows the full recognition results for both speakers with various numbers of elements discarded (the number retained are also shown for ease of comparison with later results). The word errors are summarised in Figure 2.

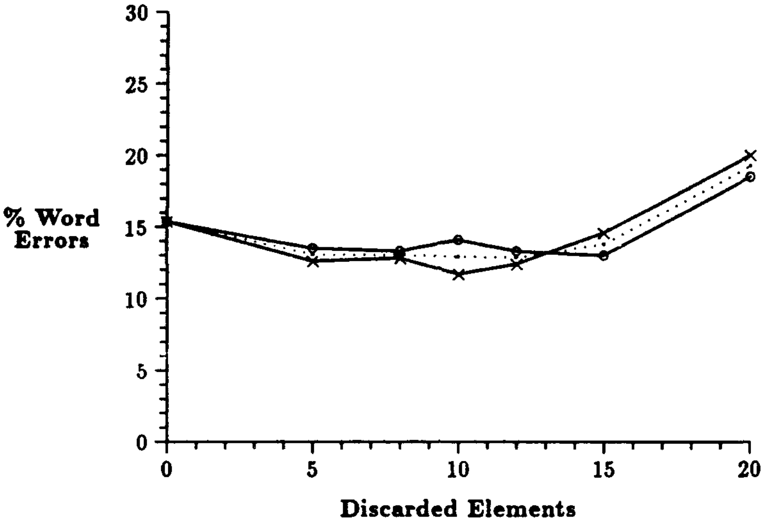

Figure 2: Word errors for various numbers of discarded elements, with data transformed using single frame transform matrices, for speakers MJR (o) and RKM (×). The dotted line represents the average over both speakers.

From these results it can be seen that not all the transformed elements are necessary in the discrimination process. Peak word recognition performance (averaged over both speakers) is obtained by discarding about ten elements, ie by retaining sixteen elements.

4.1.2 The Use of Word Transition Penalties

It was reported in [13] that significant improvements in recognition performance could be obtained by the use of suitable word transition penalties. The effect of word transition penalties on recognition performance for single frame transforms with no elements discarded (twenty six retained) is shown in Figure 3, and with ten elements discarded (sixteen retained) in Figure 4.

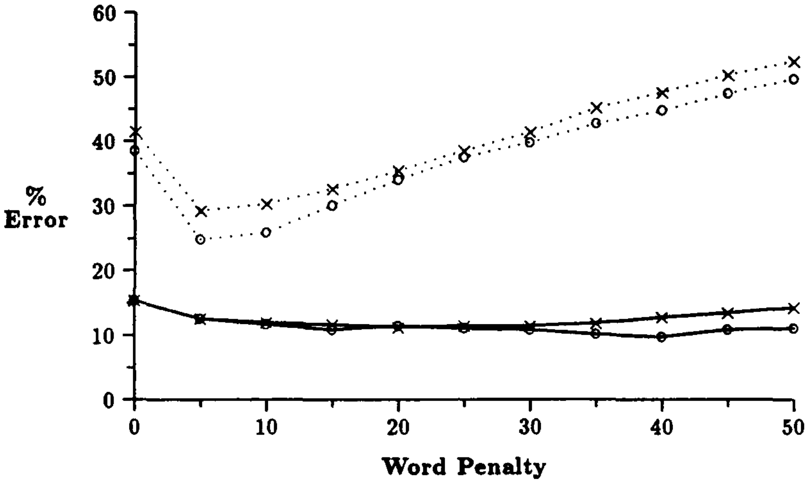

Figure 3: Word (solid line) and phoneme (dotted line) errors for various word transition penalties using single frame transform data with no elements discarded (twenty six retained) for speakers MJR (o) and RKM (×).

The behaviour is very similar in both Figures and it can be seen that very little improvement in word recognition performance is obtained from the use of word transition penalties. Some improvement in phoneme accuracy is possible by using penalties of less than about 30, whilst the best word accuracy is obtained with a penalty of 30. These results are in sharp contrast to those obtained in [13] where significant improvements in both word and phoneme accuracy were obtained by using word transition penalties.

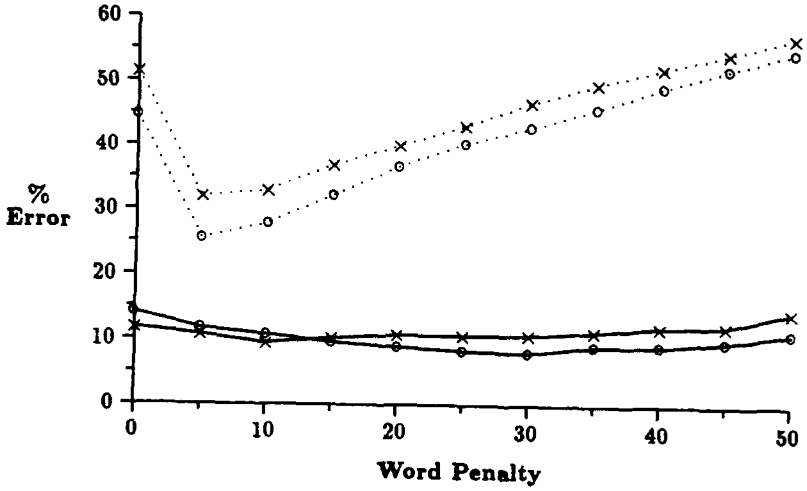

Figure 4: Word (solid line) and phoneme (dotted line) errors for various word transition penalties using single frame transform data with ten elements discarded (sixteen retained) for speakers MJR (o) and RKM (×).

4.1.3 The Use of VFR Analysis

Previous experience has shown that VFR analysis can be used to not only reduce the data rate but also to improve the recognition performance. It was therefore decided to investigate the combination of LDA and VFR analysis.

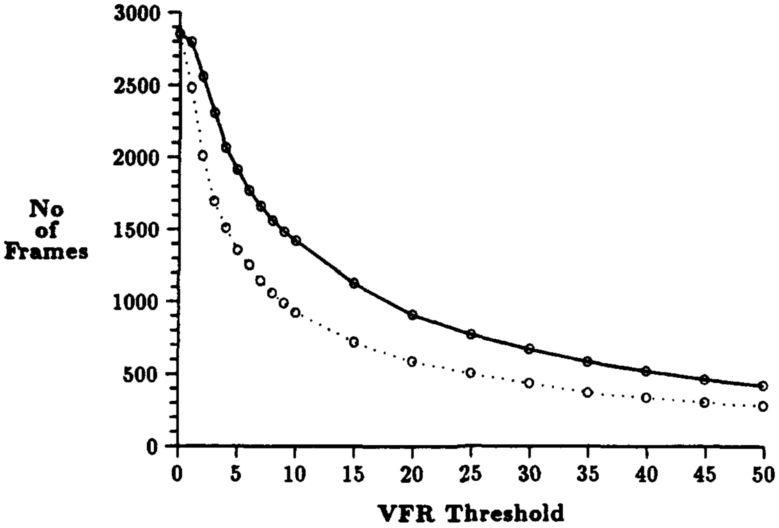

Figure 5: Number of frames processed during testing on single frame transform data for various VFR thresholds. Solid line shows effect with no elements discarded (twenty six retained) and dotted line shows the effect for ten elements discarded (sixteen retained).

In order to determine a suitable VFR threshold it was necessary to determine the effects of various VFR thresholds on a typical testing file. Figure 5 shows different VFR thresholds for single frame data with zero and ten elements discarded (twenty six and sixteen elements retained, respectively).

Previous experience had shown that a good initial value for a VFR threshold could be obtained by halving the data rate. Since it was not certain that this assumption would hold for this data, a range of values were tested. The results are shown in Table 2, with the word errors summarised in Figure 6 (in both cases the full frame rate results are included for comparison). Word transition penalties were not used in these experiments.

From these it can be seen that the use of VFR analysis results in no improvement in recognition performance when no elements are discarded. When ten

Speaker	No of elements	VFR	Phone		Word	
	Discarded/Retained	Threshold	Wrong	Errors	Wrong	Errors
MJR	0/26	0	15.1	38.5	7.2	15.4
		3	17.0	59.7	8.3	24.3
		5	16.8	53.4	7.7	22.4
		10	22.1	44.5	10.6	26.1
RKM	0/26	0	18.9	41.5	6.1	15.4
		3	20.0	61.2	6.9	19.6
		5	20.1	53.4	7.6	20.2
		10	23.7	43.1	11.9	26.9
MJR	10/16	0	15.3	44.7	6.3	14.1
		1	14.5	40.4	5.7	12.2
		2	14.5	38.8	5.9	11.7
		3	14.1	36.0	6.3	11.9
		4	16.1	36.9	6.1	11.3
		5	18.6	37.6	7.0	12.8
RKM	10/16	0	21.2	51.4	5.4	11.7
		1	21.2	49.2	6.1	13.3
		2	20.3	43.7	5.6	12.4
		3	20.9	43.8	6.5	12.4
		4	22.9	43.7	6.9	12.4
		5	22.7	41.6	8.9	17.0

Table 2: Full recognition results for single frame transform matrices with zero or ten elements discarded and VFR thresholds as shown.

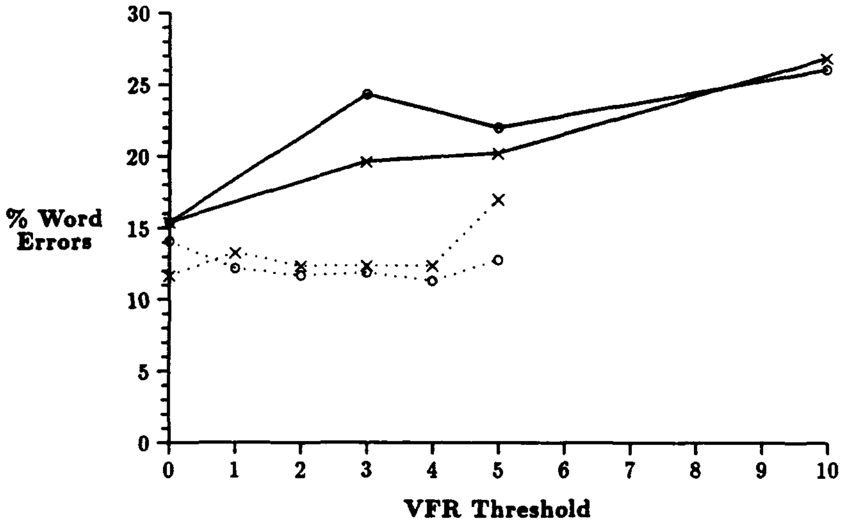

Figure 6: Word errors for single frame transform matrices with zero (solid line) or ten (dotted line) elements discarded and different VFR thresholds for speakers MJR (○) and RKM (×). No word transition penalties were used.

elements are discarded, it is possible to obtain a slight improvement for speaker MJR - with the best performance at a VFR threshold of four [2].

Various word transition penalties were then tried on the data with ten elements discarded (sixteen retained) and a VFR threshold of four. The results are not reproduced since they were very similar to those shown in Figure 4. Again there was very little improvement in word accuracy by employing word transition penalties.

[2] In fact this *is* the threshold which almost halves the original data rate.

4.2 Three Frame Transforms

For the results quoted in this section, the LDA transform matrices were created by considering three frames of input data. This resulted in output vectors containing 78 elements.

4.2.1 Varying the Numbers of Discarded/Retained Elements

Speaker	No of Elements	Phone		Word	
	Discarded/Retained	Wrong	Errors	Wrong	Errors
MJR	52/26	12.0	61.0	8.1	27.4
	56/22	10.7	41.1	6.1	16.5
	60/18	12.7	33.8	5.9	12.6
	64/14	13.9	44.2	6.1	13.0
	68/10	19.5	65.4	5.7	13.1
RKM	52/26	15.4	61.0	6.3	24.6
	56/22	15.2	36.2	4.8	11.5
	60/18	16.5	37.8	4.4	10.4
	64/14	18.8	46.4	5.2	11.1
	68/10	24.9	69.3	6.5	15.0

Table 3: Full recognition results for three frame transform matrices with various numbers of elements discarded/retained.

It was obviously impractical to use the complete output vector here so elements had to be discarded. Initially, the number to be discarded were based on experience gained with the single frame transform. The full results for phone and word errors, with various numbers of elements discarded (retained), are shown in Table 3. The word errors are summarised in Figure 7.

As in the single frame case, an improvement in word recognition performance has been obtained by discarding elements – the peak performance came from discarding about sixty elements. Hence there are eighteen elements in the output vector which correlates well with the sixteen elements for the peak performance in the single frame transform case.

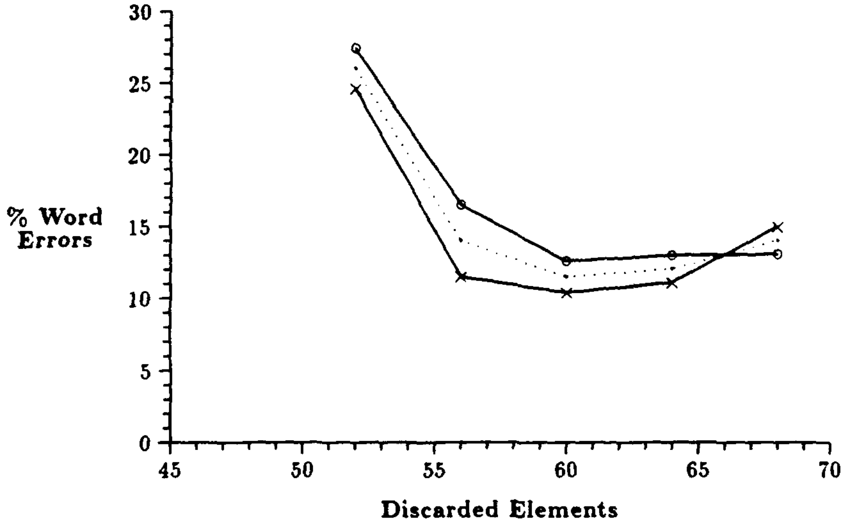

Figure 7: Word errors for various numbers of discarded elements, with data transformed using three frame transform matrices, for speakers MJR (○) and RKM (×). The dotted line represents the average over both speakers. No word transition penalties were used.

4.2.2 The Use of Word Transition Penalties

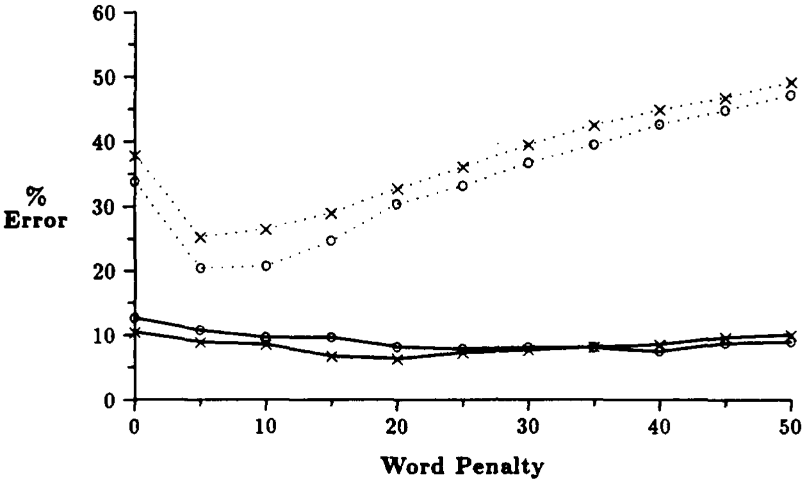

Figure 8: Word (solid line) and phoneme (dotted line) errors for various word transition penalties using three frame transform data with sixty elements discarded (eighteen retained) for speakers MJR (o) and RKM (×).

The effect of different word transition penalties on the error rates for both speakers using the three frame transform and with sixty elements discarded (eighteen retained) is shown in Figure 8. This is very similar to that shown in Figure 4 for the single frame transform data. The best word recognition performance comes from using a penalty of 20.

4.2.3 The Use of VFR Analysis

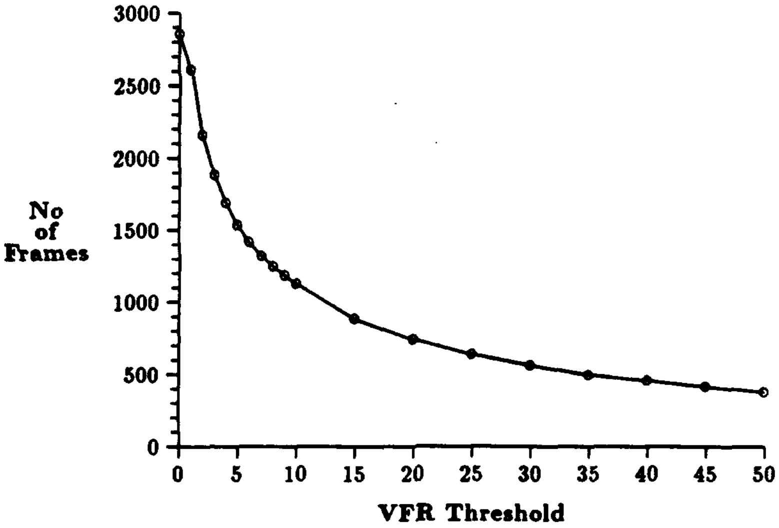

Figure 9: Number of frames processed during testing on three frame transform data, with sixty elements discarded (eighteen retained), for various VFR thresholds.

The effect of VFR analysis was investigated on the three frame transform (with sixty elements discarded) data. The effect of different VFR thresholds, on a typical testing file, is shown in Figure 9.

As before, a range of VFR thresholds were used. The results are shown in Table 4 with the word errors summarised in Figure 10.

From these it can be seen that the best results are obtained using a VFR threshold of four (which has not quite halved the data rate). As in the single frame transform case, employing word transition penalties gave no significant improvement in performance, although a penalty of 20 did result in some benefit.

Speaker	VFR	Phone		Word	
	Threshold	Wrong	Errors	Wrong	Errors
MJR	0	12.7	33.8	5.9	12.6
	1	11.7	31.4	5.7	12.2
	2	11.7	30.0	4.8	8.9
	3	11.8	29.2	5.2	9.8
	4	11.8	28.4	5.2	10.0
	5	11.6	27.0	5.4	9.8
	6	12.9	27.6	5.6	9.1
	8	15.1	28 7	8.1	13.9
RKM	0	16.5	37.8	4.4	10.4
	1	17.2	37.8	4.3	11.3
	2	16.2	37.1	4.4	10.6
	3	15.7	31.5	4.8	10.2
	4	17.6	32.1	4.6	9.4
	5	18.2	32.6	5.4	10.7
	6	18.2	32.3	6.7	12.8
	8	19.6	33.3	9.6	17.2

Table 4: Full recognition results for three frame transform matrices with sixty elements discarded (eighteen retained) and VFR thresholds as shown.

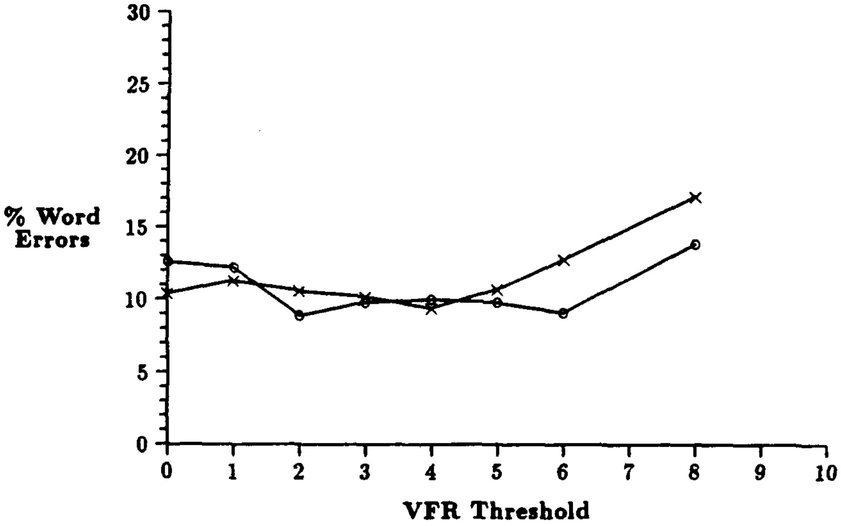

Figure 10: Word errors for three frame transform matrices with sixty elements discarded (eighteen retained) and different VFR thresholds for speakers MJR (o) and RKM (×).

5 Discussion

Prior to the use of LDA , the "best" speaker dependent recognition performance produced word errors of 10.2% for speakers MJR and RKM. This performance was obtained using 100 frames per second data, applying VFR analysis with a threshold of 500 then calculating mel frequency cepstral coefficients (no differences were employed). The recognition used a word transition penalty of 55. This result will be used as a benchmark for comparison purposes and referred to as the pre-LDA result.

Using the single frame transform matrices it was possible to match this performance. With a discard of ten (ie sixteen retained) and word penalties of 30 the word errors were 8% and 10.7% for MJR and RKM respectively. These are not significantly different ($p > 0.1$) to the pre-LDA result.

On the whole, better performance was obtained by using three frame transforms. With a discard of sixty (eighteen retained) and word penalty of 20 the word errors were 8.1% for MJR and 6.3% for RKM. This result is a significant ($p < 0.001$) improvement over the pre-LDA result for RKM but not for MJR. Using a VFR threshold of four with this data, and a word penalty of 15 produces word errors of 7% for both speakers. This is not a significant improvement, over the pre-LDA result, for either speaker.

In view of the results obtained using the three frame transform file with sixty elements discarded, the three frame transform matrices were recreated using these model files to obtain class labels for LDA. The data was transformed with these new matrices and again sixty elements were discarded. With a word penalty of 25 the word errors were 6.7% for MJR and 7.0% for RKM. These results were a significant ($p = 0.002$) improvement over the pre-LDA results for RKM, but not for MJR ($p = 0.05$)[3].

The results obtained using the recreated three frame transform matrices were not significantly ($p > 0.1$) different to those obtained with the first LDA matrices.

As stated in Section 2, a property of the LDA transform is that data which has been transformed should have diagonal between class covariance, and the within class covariance matrix should be the identity matrix. This property was checked for both three frame transforms with sixty elements discarded and it held in both cases. The two matrices for the first transform had off-diagonal elements in the range 10^{-2} to 10^{-3} whereas for the second these had decreased to 10^{-9} to 10^{-12}, which is much more satisfactory.

[3] The different significance levels obtained for the same reduction in overall word errors (10% to 7% here and above) for RKM are a good illustration of the importance of considering the detailed error pattern, cf [2].

6 Conclusions

These are preliminary conclusions based on a small set of speaker dependent experiments. See [10] for a more general report based on speaker independent experiments.

The conclusions for this speaker dependent database are:–

- LDA can be used to significantly improve the performance of the speaker dependent *ARM* system by using three frame transform matrices and suitable word transition penalties.
- The best pre-LDA result can be matched using single frame LDA transform matrices and suitable word transition penalties.
- Some improvement in performance can be achieved by using VFR analysis on LDA transformed data.
- If VFR analysis is used, halving the data provides a good initial guess for a suitable threshold to use.
- Word transition penalties can be used to achieve some improvement in recognition accuracy, but this improvement is not as marked as in previous cases ([13]).

References

[1] J S Bridle, M D Brown and R M Chamberlain, "A One-Pass Algorithm for Connected Word Recognition", Froceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Paris, pp899–902, 1982.

[2] L Gillick and S J Cox, "Some Statistical Issues in the Comparison of Speech Recognition Algorithms", Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, Glasgow, 23–26 May, pp532–535, 1989.

[3] J N Holmes, "The JSRU Channel Vocoder", IEE Proceedings, vol 127, Part F, number 1, pp 53–60, February 1980.

[4] M J Hunt, "An Introduction to Linear Discriminant Analysis", JSRU Research Report no 1007, 1978.

[5] M J Hunt and C Leféb vre, "Distance Measures for Speech Recognition", National Aeronautical Establishment, Canada, NAE-AN-57, NRR No. 30144, March 1989.

[6] K-F Lee, "Large Vocabulary Speaker-Independent Continuous Speech Recognition: the SPHINX System", PhD Thesis, Carnegie Mellon University, 1988.

[7] D S Pallett, W M Fisher, and J G Fiscus "Tools for the Analysis of Benchmark Speech Recognition Tests", Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Albuquerque, New Mexico, pp97–100, 1990.

[8] D B Paul, "The Lincoln Robust Continuous Speech Recogniser", ICASSP 89, Glasgow, Scotland, pp449–452, 1989.

[9] S M Peeling and K M Ponting, "Speaker Dependent Speech Recognition Experiments Using Alternative Front Ends With Variable Frame Rate Analysis", RSRE Memo 4389, 1990.

[10] S M Peeling and K M Ponting, "The Use of Linear Discriminant Analysis in the *ARM* Continuous Speech Recognition System", RSRE Memo 4512, 1992.

[11] L C W Pols, "Spectral Analysis and Identification of Dutch Vowels in Monosyllabic Words", Academische Pers B.V., Amsterdam, 1977.

[12] K M Ponting and S M Peeling, "Experiments in Variable Frame Rate Analysis for Speech Recognition", RSRE Memo 4330, 1989.

[13] K M Ponting and S M Peeling, "Word Transition Penalties in the *ARM* Continuous Speech Recognition System", RSRE Memo 4362, 1990.

[14] K M Ponting and M J Russell, "The ARM Project: Automatic Recognition of Spoken Airborne Reconnaissance Reports", Proceedings of 'Military and Government Speech Tech 89', Arlington VA, pp223–227, 13–15 November 1989.

[15] M J Russell, K M Ponting, S M Peeling, S R Browning, J S Bridle, R K Moore,I Galiano and P Howell, "The ARM Continuous Speech Recognition System" Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Albuquerque, New Mexico, pp69–72, 1990.

REPORT DOCUMENTATION PAGE

DRIC Reference Number (if known) ....

Overall security classification of sheet ..................................UNCLASSIFIED..

(As far as possible this sheet should contain only unclassified information. If it is necessary to enter classified information, the field concerned must be marked to indicate the classification eg (R), (C) or (S).

Originators Reference/Report No.
MEMO 4511

Month
DECEMBER

Year
1991

Originators Name and Location
RSRE, St Andrews Road
Malvern, Worcs WR14 3PS

Monitoring Agency Name and Location

Title

PRELIMINARY RESULTS ON THE USE OF LINEAR DISCRIMINANT ANALYSIS IN THE *ARM* CONTINUOUS SPEECH RECOGNITION SYSTEM

Report Security Classification
UNCLASSIFIED

Title Classification (U, R, C or S)
U

Foreign Language Title (in the case of translations)

Conference Details

Agency Reference

Contract Number and Period

Project Number

Other References

Authors
PEELING, S M; PONTING, K M

Pagination and Ref
20

Abstract

Linear discriminant analysis is used to generate speech data transformations. This transformed data is then used within the *ARM* continuous speech recognition system. Preliminary results are presented from experiments using transformed data alone and also in conjunction with one, or both, of word transition penalties and variable frame rate analysis. Speaker dependent results are reported which are significantly better than the best obtained previously.

Abstract Classification (U,R,C or S)
U

Descriptors

Distribution Statement (Enter any limitations on the distribution of the document)
UNLIMITED

880/48

INTENTIONALLY BLANK